14

Bosses and Workers, Landlords and Peasants, and other Principal-Agent Models

14.1 Introduction to the Principal-Agent Model

In the *principal-agent model* the payoff to the *principal* depends on an action taken by the *agent*. The principal cannot contract for the action, but can compensate the agent based on some observable signal that is *correlated* with the action. The principal is *first mover*, as in a Stackelberg duopoly model (§3.5) and chooses an incentive scheme for paying the agent that depends on the observed signal. The principal offers this incentive scheme, as a contract to the agent. The agent then (a) determines the optimal action to take, given the incentives, and then (b) decides whether or not to accept the contract, based on the expected payment and the subjective cost of performing the action. Upon accepting the contract, the agent performs the optimal action, the principal observes the signal correlated with the action, pays the agent according to the incentive scheme, and receives a payoff dependent upon the signal.[1]

For example, suppose the principal is an employer. The agent is then a worker, the worker's action is a level of work effort, the signal may be the employer's profits or a supervisor's report on the worker's behavior, the incentive scheme is a wage that depends on the signal, and the employer's payoff is expected profits, which depends on the worker's effort level. For another example, suppose the principal is an insurance company. The agent may be the owner of a warehouse who purchases fire insurance from the company, the owner's action may be a degree of care in avoiding a fire, the incentive scheme may be a percentage of co-insurance

[1]Notice that the *incentive scheme* is contractible, even though the agent's *action* is not. If the principal cannot precommit to pay the incentives, or if the agents' actions can affect the incentive scheme itself, then some other model must be used (e.g., a bargaining model).

VISIT…

for the owner and the signal may be an observed level of fire damage.[2] You have seen principal-agent problems before, including *Contingent Renewal Contracts* (§6.14), which discusses a variety of models, *Markets as Disciplining Devices: Allied Widgets* (§3.18), in which the principal is the firm's owner and the agent is the firm's manager, and *I'd Rather Switch than Fight* (§6.16), in which the principal is the consumer and the agent is the supplier.[3] Here we study such problems more systematically and in greater depth.

14.2 Labor Discipline with Monitoring

An employer hires supervisors to watch the workers, docking the pay of any worker who is caught shirking. Worker effort takes the form of working a certain fraction e of the time, so if there are N workers and each worker works at effort level e for one hour, then total labor supplied is eN. All workers have the same utility function $u(w, e) = (1-e)w$ where w is the wage rate and e is the effort level. The worker is normally paid w, but is paid $z < w$ if caught shirking.

Suppose that a worker who chooses effort level e is caught shirking with probability $p(e) = 1 - e$, so the harder the worker works, the lower the probability of being caught shirking. The game tree for this problem in depicted in Figure 14.1

(a) Show that the payoff to a worker who chooses effort level e is $w(1-e)e + z(1-e)^2$;

(b) Show that if the employer offers wage $w > 2z$, the worker's best

[2] In case you've never heard of co-insurance, it goes like this: if the warehouse would cost $1,000,000 to replace and the co-insurance rate is 20%, then if the warehouse were totally destroyed by fire, the insurance company would only pay $800,000. Note that if the market value of the warehouse dipped below $800,000 for some reason, the owner would pray for a fire. Or perhaps more than pray, as in the following anecdote: Manny is taking his morning constitutional, and bumps into his old friend Sol, who exclaims "Manny, I heard about the terrible fire at your warehouse. I'm so sorry for you." "Shhhh!" replies Manny, "It's tomorrow!" As a gruesome example of the same phenomenon, a few years back the bottom dropped out of the race horse market, so the insurance value of a thoroughbred horse far exceeded it market value. Many owners saw to it that their horses suffered 'accidental' broken legs, recovering their losses at the expense of the insurance companies (and the horses).

[3] Arguably, principal-agent models best describe the most important transactions in a market economy, involving capital, labor, and product markets alike. See §6.14 for an analysis of the broad implications of this fact.

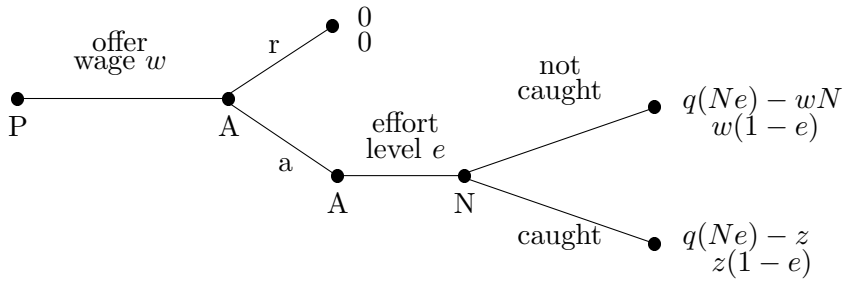

Figure 14.1: Labor Discipline with Monitoring

response is to choose

$$e(w) = \frac{w - 2z}{2(w - z)}.$$

Show that this worker *best response function* is increasing and concave, and hence is as depicted in Figure 14.2.

(c) If the employer chooses w and N to maximize profits, show that the choice of w in fact maximizes $e(w)/w$, the amount of effort per dollar of wages, which is the slope of the employer iso-cost line in Figure 14.2.

(d) Show that Nash equilibrium (w^*, e^*) satisfies the *Solow condition*

$$e'(w^*) = \frac{e(w^*)}{w^*}.$$

This is where the employer iso-cost line is tangent to the worker's best response function at (w^*, e^*) in Figure 14.2.

(e) Show that

$$w^* = (2 + \sqrt{2})z \approx 3.41z, \qquad e^* = \frac{1}{\sqrt{2}(1 + \sqrt{2})} \approx 0.29.$$

(f) Suppose the worker's reservation utility is $z_0 > 0$, so the worker must be offered expected utility z_0 to agree to come to work. Show that the employer will set $z = (1 + \sqrt{2})/(3 + 2\sqrt{2})z_0 \approx 0.83z_0$.

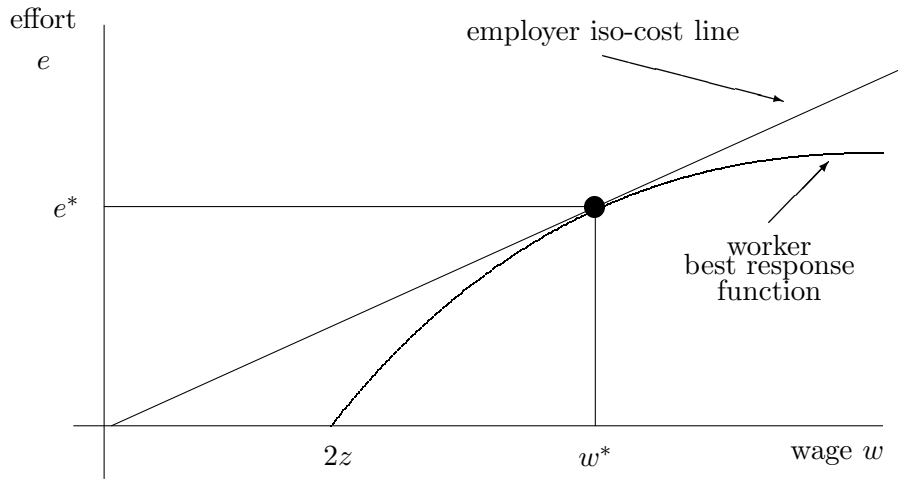

Figure 14.2: Equilibrium in the Labor Discipline Model

14.3 Labor as Gift Exchange

In a famous paper, "Labor Contracts as Partial Gift Exchange," George Akerlof (1982) suggested that when employers pay workers more than absolutely necessary, workers often reciprocate by working harder or more carefully than they otherwise would. He called this *gift exchange*. Economists generally ignored this possibility until Ernst Fehr, working at the University of Zürich, showed that laboratory experiments strongly confirm the existence of Akerlof's gift exchange phenomenon (see also Fehr's related work on reciprocity in §11.4.2 and §11.4.1). Specifically, whereas under competitive experimental conditions without the opportunity for buyers and sellers to reciprocate, prices and quantities converge to the competitive equilibrium, when reciprocation can be expressed, many buyers and sellers behave reciprocally, driving a competitive experimental market permanently away from the competitive outcome (Fehr, Gächter, Kirchler, and Weichbold 1998, Fehr, Kirchsteiger, and Riedl 1998).

This section analyzes a simple model of gift exchange in labor markets. A firm hires N identical workers, each of whom supplies effort level $e(w - z)$ where $e(\cdot)$ is increasing and concave, w is the wage, and z is a benchmark wage such that $w > z$ indicates that the employer is being generous, and conversely $w < z$ indicates that the boss is a skinflint.

The firm's revenue is an increasing and concave function $f(eN)$ of total amount of effort supplied by the N workers, so the firm's net profit is given by

$$\pi(w, N) = f(e(w - z)N) - wN.$$

Suppose that the firm chooses w and N to maximize profits.

(a) Show that the *Solow condition*, $de/dw = e/w$, holds (Solow 1979);

(b) Show that the second order condition for a profit maximum is satisfied;

(c) Show that the following comparative static results hold:

$$\frac{de}{dz} > 0; \quad \frac{d\pi}{dz} < 0 \quad \frac{dw}{dz} > 0.$$

This analysis show that when workers' reference wage z increases, employers respond by *increasing* the gift component $w - z$ in the wage, leading to higher effort but lower profits.

14.4 Labor Discipline with Profit Signaling

An employer hires a worker to do a job. There are two possible levels of profits for the employer, high (π_H) and low ($\pi_L < \pi_H$). The worker can affect the probability of high profits by choosing to work with either high or low effort. With high effort the probability of high profits is p_h, and with low effort the probability of high profits p_l, where $0 < p_l < p_h < 1$.

If the employer could see the worker's choice of effort, he could simply write a contract for high effort, but he cannot. The only way he can induce A to work hard is to offer the proper *incentive contract*: pay a wage w_H if profits are high, and $w_L < w_H$ if profits are low.

How should the employer choose the incentives w_L and w_H to maximize expected profits? The game tree for this situation is shown in Figure 14.3, where we assume the utility of the wage is $u(w)$, the cost of high effort to the worker is d_h and the cost of low effort is $d_l < d_h$. By working hard, the worker faces a lottery with payoffs $(u(w_H) - d_h, u(w_L) - d_h)$ with probabilities $(p_h, 1 - p_h)$, the expected value of which is

$$p_h(u(w_H) - d_h) + (1 - p_h)(u(w_L) - d_h) = p_h u(w_H) + (1 - p_h)u(w_L) - d_h.$$

With low effort, the worker faces a lottery with payoffs $(u(w_H) - d_l, u(w_L) - d_l)$ with probabilities $(p_l, 1 - p_l)$, the expected value of which

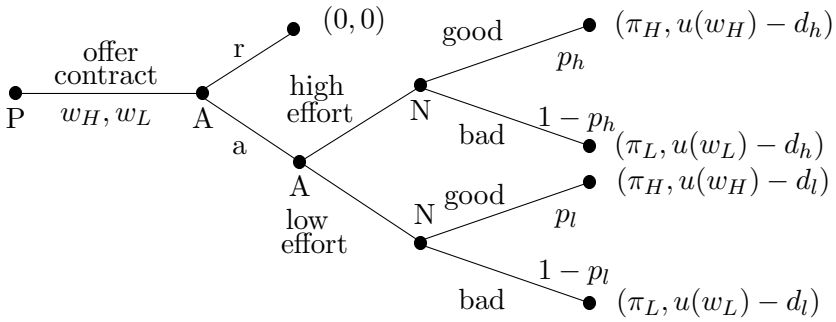

Figure 14.3: Labor Incentives

is

$$p_l(u(w_H) - d_l) + (1 - p_l)(u(w_L) - d_l) = p_l u(w_H) + (1 - p_l)u(w_L) - d_l.$$

Thus the worker will choose high effort over low effort only if the first of these expressions is at least as great as the second. For this to be the case, we must have $p_h u(w_H)+(1-p_h)u(w_L)-d_h > p_l u(w_H)+(1-p_l)u(w_L)-d_l$, which simplifies to

$$(p_h - p_l)(u(w_H) - u(w_L)) \geq d_h - d_l. \tag{14.1}$$

This is called the *incentive compatibility constraint* for eliciting high effort.

The incentive compatibility constraint (14.1) shows that the larger the cost to the worker of supplying high effort $(d_h - d_l)$, the larger the wage differential $(w_H - w_L)$ must be. Also, the smaller the difference in the probability of affecting the outcome $(p_h - p_l)$, the larger the wage differential must be. This makes sense, since if $p_h - p_l$ is very small, the worker's payoff is not very different whether he chooses high or low effort, unless his wage differential is very large.

Now suppose the worker's next best job prospect has expected value z. Then to get the worker to take the job, the employer must offer him at least z. This the following *participation constraint*:

$$p_h u(w_H) + (1 - p_h)u(w_L) - d_h \geq z, \tag{14.2}$$

assuming the principal wants the worker to work hard.

The expected profit of the employer, assuming the worker works hard, is given by

$$p_h(\pi_H - w_H) + (1 - p_h)(\pi_L - w_L). \tag{14.3}$$

It's clear that to minimize the expected wage bill, the employer should choose w_H and w_L so that equation (14.2) is satisfied as an equality, since this minimizes the expected wage bill. Also, he should choose w_H and w_L so that equations (14.1) and (14.2) are satisfied as equalities. This is illustrated in Figure 14.4.

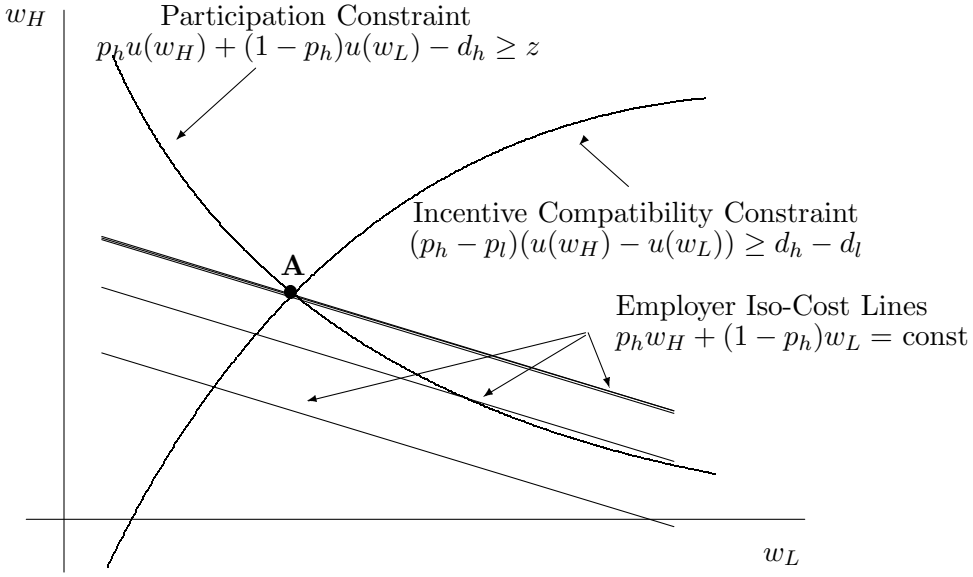

Figure 14.4: Minimizing the Cost of Inducing an Action, given Participation and Incentive Compatibility Constraints

Using this figure, prove the following:

(a) Employer iso-cost lines are of the form $p_h w_H + (1-p_h)w_L = \text{const}$;

(b) The participation constraint is decreasing and convex;

(c) The incentive compatibility constraint is increasing and cuts the w_L-axis for some $w_L > 0$. If the agent is weakly decreasingly risk averse (i.e., if $u''' > 0$), then the incentive compatibility constraint is concave.

(d) If the agent is strictly risk averse (§rbsufsec), the slope of the iso-cost lines is less than the slope of the participation constraint at its intersection A with the incentive compatibility constraint;

It follows that the solution is at **A** in Figure 14.4.

14.4.1 Properties of the Labor Discipline Model

The above exercise shows that the unique Nash equilibrium in the model is the solution to the two equations $p_h u(w_H) + (1 - p_h)u(w_L) - d_h = z$ and $(p_h - p_l)(u(w_H) - u(w_L)) = d_h - d_l$. Solving simultaneously, we get

$$u(w_L) = z + \frac{p_h d_l - p_l d_h}{p_h - p_l} \tag{14.4}$$

$$u(w_H) = u(w_L) + \frac{d_h - d_l}{p_h - p_l}. \tag{14.5}$$

Notice that the worker in this case is no better off than if he did not take the job—this is a one-shot incentive model, not a realistic labor discipline model, in which the employer pays the worker a wage higher than his fallback and hires the worker for an indefinite number of periods, using the threat of dismissal to elicit effort from the worker (§6.15).

As we might expect, if z rises, so do the two wage rates w_L and w_H. If d_h rises, you can check that w_H rises and w_L falls. Similar results hold when p_h and p_l vary.

Now that we know the cost to the principal of inducing the agent to take each of the two actions, we can determine which action the principal should have the agent choose. If H and L are the expected profits in the good and bad states, respectively, then the return $\pi(a)$ for inducing the agent to take action $a = h, l$ is given by

$$\pi(h) = Hp_h + L(1 - p_h) - \mathbf{E}_h w, \qquad \pi(l) = Hp_l + L(1 - p_l) - \mathbf{E}_l w,$$

where $\mathbf{E}_h w$ and $\mathbf{E}_l w$ are the expected wage payments if the agent takes actions h and l respectively; i.e., $\mathbf{E}_h w = p_h w_H + (1 - p_h)w_L$ and $\mathbf{E}_l w = p_l w_H + (1 - p_l)w_L$.

Is it worth inducing the worker to choose high effort? For low effort only the participation constraint $u(w_l) = d_l + z$ must hold, where w_l is the wage paid independent of whether profits are H or L, with expected profit $p_l H + (1 - p_l)L - w_l$. Choose the incentive wage if and only if $p_h(H - w_H) + (1 - p_h)(L - w_L) \geq p_l H + (1 - p_l)L - w_l$. This can be written

$$(p_h - p_l)(H - L) \geq p_h w_H + (1 - p_h)w_L - w_l. \tag{14.6}$$

We have seen that in general, if the worker is risk neutral, and it is worth exerting high effort, then the optimum is to make the principal the fixed claimant and the agent the residual claimant (§14.7). To see this for the current example, we can let $u(w) = w$ (why is this okay?).

The participation constraint is $p_h u(w_H) + (1-p_h)u(w_L) = p_h w_H + (1-p_h)w_L = z + d_h$, the employer's profit is then $A = p_h H + (1-p_h)L - (z+d_h)$. Suppose we give A to the employer as a fixed payment, and let $w_H = H - A$, $w_L = L - A$. Then the participation constraint holds, since

$$p_h w_H + (1-p_h)w_L = p_h H + (1-p_h)L - A = z + d_h.$$

Since high effort is superior to low effort for the employer, equation (14.6) must hold, giving

$$\begin{aligned}(p_h - p_l)(H-L) &\geq p_h w_H + (1-p_h)w_L - (z+d_l)\\ &= z + d_h - (z+d_l) = d_h - d_l.\end{aligned}$$

But then

$$w_H - w_L = H - L \geq \frac{d_h - d_l}{p_h - p_l},$$

which says that the incentive compatibility constraint is satisfied.

Figure 14.5 is a graphical representation of the principal's problem. Note that in this case there are many profit-maximizing contracts—any point on the heavy solid line in the figure maximizes profits.

14.5 Peasants and Landlords

A landlord hires a peasant to plow, sow, and reap a corn field. The landlord's profit is H if the crop is good, and $L < H$ if the crop is poor. The peasant can either work at high effort h or low effort l, and the probability p_h of a good crop when he exerts high effort is greater than the probability p_l of a good crop when he expends low effort, with $0 < p_l < p_h < 1$. The landowner cannot observe the peasant's effort.

Suppose the peasant's utility function when the wage is w is given by $u(w) - d_h$ with high effort and $u(w) - d_l$ with low effort. We assume $d_h > d_l$, so unless given some inducement, the peasant will not work hard, and $u' > 0, u'' < 0$, so the peasant has diminishing marginal utility of the wage. The peasants fallback utility is z.

To induce the peasant to work hard, the landlord chooses pair of wages w_H and w_L, and pays the peasant w_H if profit is H, and w_L if profit is L. This is called an *incentive wage*.

(a) What should the landlord pay the peasant if he wants minimize the expected wage $Ew = p_h w_H + (1-p_h)w_L$, subject to eliciting high effort?

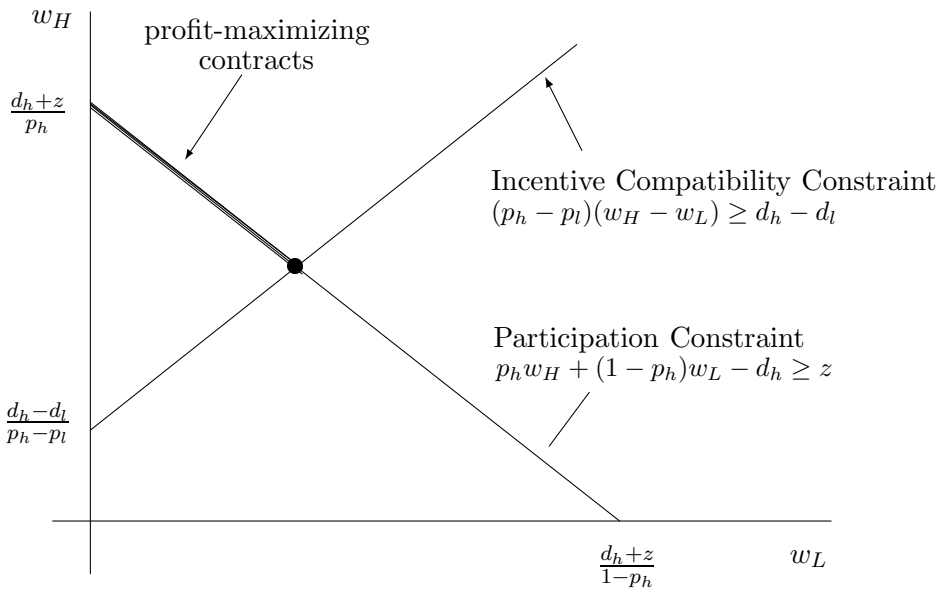

Figure 14.5: The Principal's Problem when the Agent is Risk Neutral

(b) Using the appropriate Kuhn-Tucker conditions, prove that both the PC and the ICC must hold as equalities.

(c) Prove that the optimal incentive wage for the landlord is given by

$$u(w_L) = d_h - p_h(d_h - d_l)/(p_h - p_l) + z$$
$$u(w_H) = d_h + (1 - p_h)(d_h - d_l)/(p_h - p_l) + z$$

(d) Suppose now that the peasant is risk neutral but the landlord is risk averse (§rbsufsec). Show that if high effort produces a surplus over low effort, the landlord's profit-maximizing solution involves getting a fixed rent from the peasant, who becomes the residual claimant, bearing all the risk and taking all the profits and losses.

14.6 Mr. Smith's Car Insurance

Mr. Smith wants to buy theft insurance for his car, which is worth \$1200. If Smith is careful, the probability of theft is 50%, and if he is careless, the probability of theft is 75%. Smith is risk averse with utility function $u(x) = \ln(x+1)$, and the disutility of being careful is $\epsilon > 0$. To keep the

problem simple, suppose the car lasts one period, and is either stolen or not stolen at the beginning of the period.

(a) What is the expected utility of the car to Mr. Smith? For what range of values of ϵ is it worthwhile for Smith to be careful without insurance?

(b) Suppose Mr. Smith buys an insurance contract for premium x, that pays \$1200 if his car is stolen. Show that Smith will not be careful, and find the premium to the insurance company, assuming the company is risk neutral and competition forces the insurance policy to be a fair lottery.

(c) What is the expected utility of the car plus the insurance policy for Smith?

(d) Show that it is worthwhile for Mr. Smith to buy the insurance.

(e) Suppose Mr. Smith would be careful without insurance, and the insurance company can charge a deductible of z, so if the car is stolen, the company pays \$1200 $- z$. Find the value of z that maximizes the expected utility of the car plus insurance policy for Smith.

(f) Show that Smith is now careful with insurance.

Notice that this problem, while clearly in the spirit of a principal-agent model, differs from the generic case in one crucial way: the principal rather then the agent is forced to his reservation position (a zero-profit condition) rather than the agent, since we assume a competitive market for insurance.

14.7 A Generic One-Shot Principal Agent Model

Principal-agent models are variations of the following scenario. The agent has a set $A = \{a_1, \ldots, a_n\}$ of available actions. The principal receives one of a set $S = \{s_1, \ldots, s_m\}$ of signals that depend on the action chosen by the agent. Let $p_{ij} = \mathrm{P}[s_j|a_i]$ be the conditional probability of signal s_j when the agent takes action a_i. Note that we must have

$$\sum_{j=1}^{m} p_{ij} = 1 \qquad \text{for } i = 1, \ldots, n.$$

The agent has utility function $u(w, a)$, where w is money and $a \in A$ is the action he performs, and has a reservation utility is u_0, which is

the minimum expected utility that would induce him to work for the principal. The principal has payoff $\pi(a) - w$ for $a \in A$.[4]

To maximize his payoff, the principal must determine the payoff associated with getting the agent to perform each action $a_i \in A$, and then choose the largest of the n payoffs. Thus for each $i = 1, \ldots, n$ the principal must find incentives $w_{ij}, j = 1, \ldots, m$ such that the principal agrees to pay the agent w_{ij} if he observes the signal s_j, and the incentive scheme $w_i = \{w_{i1}, \ldots, w_{im}\}$ induces the agent to choose a_i. The principal must then choose the index i that maximizes his return

$$\pi(a_i) - \mathbf{E}_i w_i = \pi(a_i) - \sum_{j=1}^{m} p_{ij} w_{ij},$$

where $\mathbf{E}_i$ is expectation with respect to the probabilities that obtain when the agent chooses a_i.

There are two constraints to the problem. Suppose the principal wants to induce the agent to choose action a_k. Then we have:

Participation Constraint. The expected utility to the agent must be enough to induce the agent to participate: $\mathbf{E}_k u(w_k, a_k) \geq u_0$, or

$$\sum_{j=1}^{m} p_{kj} u(w_{kj}, a_k) \geq u_0.$$

Incentive Compatibility Constraint. The chosen action a_k must maximize the payoff to the agent, among all actions $a_i \in A$, so $\mathbf{E}_k u(w_k, a_k) \geq \mathbf{E}_i u(w_k, a_i)$ for $i = 1, \ldots, n$, or

$$\sum_{j=1}^{m} p_{kj} u(w_{kj}, a_k) \geq \sum_{j=1}^{m} p_{ij} u(w_{kj}, a_i) \qquad \text{for } i = 1, \ldots, n.$$

Having determined the minimum expected cost $\mathbf{E}_k w_k$ of inducing the agent to choose action a_k, the principal chooses the index k to maximize $\pi(a_k) - \mathbf{E}_k w_k$.

Prove the following

Theorem 14.1. Fundamental Theorem of the One-Shot Principal-Agent Model. *Every solution to the principal-agent problem has the following characteristics:*

[4]We assume $\pi(a)$ has a stochastic component, so the principal cannot infer a from $\pi(a)$.

(a) The agent is indifferent between participating and taking the reservation utility u_0;

(b) If the agent is strictly risk averse and the principal's optimal action a_k is not the agent's most preferred action, then

(i) At least one of the incentive compatibility constraints holds as an equality;

(ii) The principal's payoff is strictly lower than it would be if he could write an enforceable contract for the delivery of action a_k.

(c) If the agent is risk neutral, and if the return $\pi(a)$ is transferable (i.e., it can be assigned to the agent), the contract that maximizes the principal's payoff is that which gives him a fixed payment and makes the agent the residual claimant on $\pi(a)$.

The last finding is one of the most important in all of game theory, I believe. It says that individual agents (e.g., farmers, workers, managers) should always be residual claimants on their projects, as long as they are risk neutral and the return $\pi(a)$ to their actions is transferable. In practice, however, such agents are normally *not* residual claimants on their projects—managers may have salaries plus stock options and farmers may have sharecropping contracts, but these contractual forms do not render the principals fixed claimants. What is wrong? It could be that principals (landlords, stockholders) do not choose optimal contracts. More likely, however, agents are emphrisk averse, so that they would not want to be residual claimants, or they are *credit constrained*, so they cannot credibly promise to make good on the project's losses in bad times. Moreover, $\pi(a)$ is often *not* transferable (e.g., in a large work team, it is difficult to isolate the contribution of a single worker). At any rate, the theory of who controls productive activity is brightly illuminated by the principal-agent model.